# A Higher Level of Performance

## How SAS software exploits the newest technologies of the Pentium III Xeon processor-based server for enterprise decision support

The evolution of the Intel architecture has improved various aspects of the decision-support server, leading to systems that are achieving a higher level of performance in many areas. These improvements are important to SAS Institute customers because they improve the ability of these customers to extract knowledge from their data, resulting in competitive advantage and positive impact on the bottom line.

The new Pentium III Xeon processor offers several significant performance improvements: increased processor speed, higher system bus speed, larger L2 caches, ESMA, increased SMP processor limits, and Streaming SIMD Extensions. These attributes make it an important consideration for IT managers with systems running enterprise-level data warehouses and performing computational-intensive data mining and customer relationship management (CRM) tasks, among other applications.

Let's take a closer look at how SAS Institute, the leader in enterprise decision support, is helping its customers leverage the technologies of the Pentium III Xeon processor-based servers:
**System speed** – For companies implementing enterprise resource planning (ERP) and other heavy transactional processing systems, as well as enterprise data warehouses and CRM solutions – like those from SAS Institute – that pull data out of these systems and make it useful for business decision-making, the speed of the microprocessor is key. The new Pentium III Xeon processor provides a fast 100 MHz system bus with the processor running up to 550Mhz, alleviating system bottlenecks and optimizing I/O and memory performance for a balanced platform that can support critical business applications.
**Larger cache memory** – The new processor features a large L2 cache (up to 2M). Cache is a special type of memory that stores frequently accessed memory locations and their addresses. Since cache memory is much faster than normal memory (RAM), a larger cache can hold more information and significantly speed up operations. Both the higher system speed and the

increased cache mean better performance for data warehousing, data mining, and transactional applications that involve database access.

**Monterey, operating system choice** – While Windows NT is increasingly scalable and reliable for mission-critical applications, the Pentium III Xeon processor will also support SCO UnixWare, giving the customer a choice in operating environment. Project Monterey, announced in October 1998, represents an opportunity for unification of the UNIX marketplace around a single implementation which is supported by both the volume leader of UNIX on Intel servers (SCO) and enterprise technology leaders (IBM & Sequent). SAS® software supports Windows NT and SCO UnixWare running on the Intel architecture, as well as all major UNIX operating environments. In fact, SAS software supports more than 30 different operating environments from the desktop to the mainframe, giving customers *maximum choice* in scaling enterprise applications and leveraging existing platform investments.

**Extended server memory architecture (ESMA)** – In the Pentium III Xeon processor, Intel has added support for much larger amounts of memory through an architecture called ESMA (Extended Server Memory Architecture). ESMA provides a fundamental performance benefit: the ability to hold all or a large portion of a customer's enterprise database in main memory (RAM) as opposed to repeatedly moving that data to and from the disk susbsystem. Disk transfers are many times slower than data transfers between the processor and main memory. ESMA allows applications to take advantage of main memory beyond the Windows-NT limit of four gigabytes. In currently available server hardware, this means availability of main memory of from eight to 32 gigabytes.

Beginning with Version 8 of SAS software (scheduled for release later this year) running on Windows NT, ESMA capabilities are available to users of the SAS® System's suite of data warehousing and decision support software. Through SAS Institute's support for this architecture, SAS applications can make direct use of memory beyond the four-gigabyte limit to greatly improve access for various types of data processing operations. Initial tests have shown significant performance improvements for some types of operations that make use of ESMA, particularly those that require access to large amounts of disk-based data. In these cases, when much or all of the data can reside in memory, as compared with relatively slow disk drives, the central processing power can be more effectively used.

In addition to lowering or eliminating the overhead of waiting for disk systems to handle large data requests, ESMA has several other benefits when it is used effectively:

- when information is retrieved through ESMA, it can dramatically decrease disk channel traffic, resulting in greater performance of other operations that may be occurring on a server system.
- since ESMA can in some cases be immune to disk bottlenecks, main SAS software processing throughput can remain largely unaffected by other server activities, as long as central processing power is available.
- SAS software ESMA operations bypass the Windows NT disk cache, leaving it available to handle other application requirements on a server system.
- the amount of memory available to SAS software processing is limited only by hardware availability. This can be far greater than typical applications can commonly access in Windows NT, which is limited to three gigabytes per application. When these very large amounts of memory are available, very impressive performance improvements can result.

***[For more on the effects of ESMA on SAS software performance, SEE CHART 1]***

**Streaming SIMD (single instruction, multiple data) extensions** – SAS software also takes advantage of a new feature of the Pentium III Xeon processor called Streaming SIMD extensions. Streaming SIMD extensions, formerly known as Katmai new instructions, allow the processor to work on more data than just 32 bits at a time, which is typically what earlier processors supported. This feature allows up to 4 separate pieces of data to be operated upon simultaneously.

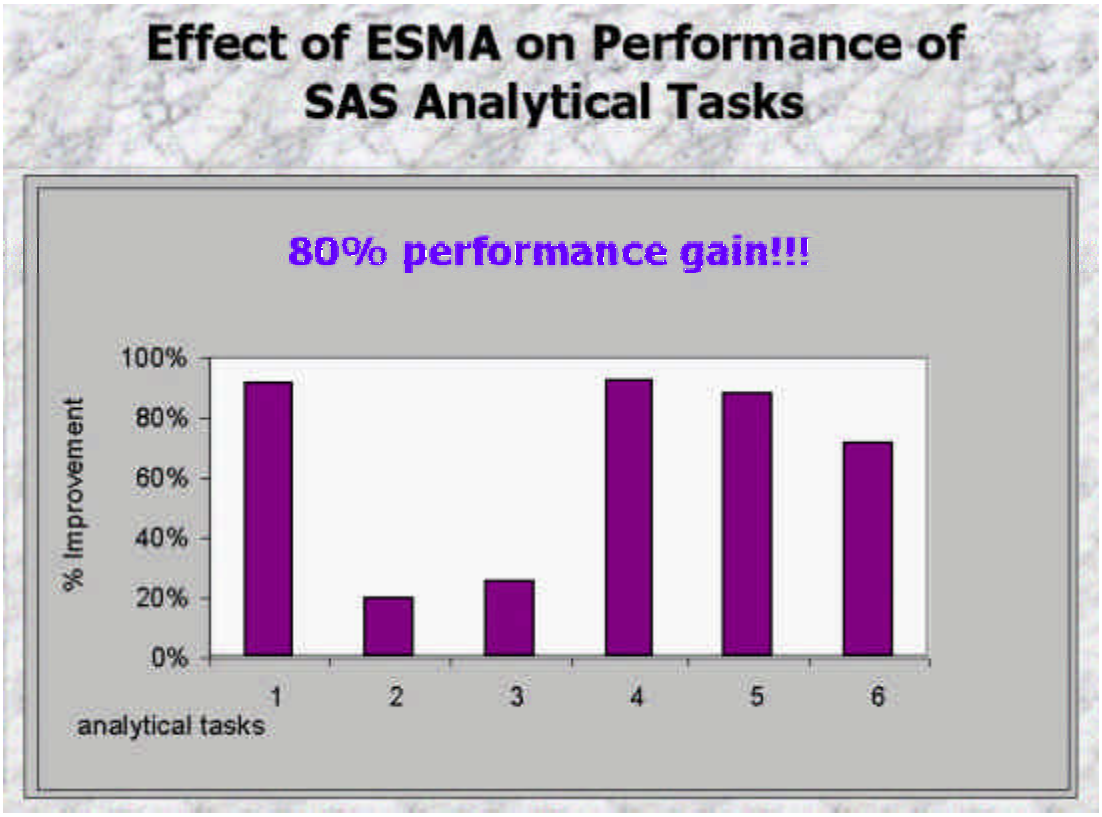
SAS software takes advantage of Streaming SIMD extensions for certain involved calculations used in analytic modeling techniques, resulting in a performance benefit for these types of operations. Specifically, it maximizes processor utilization by pre-fetching data in advance of its actual use.

***[For more on the effects of Streaming SIMD extensions on SAS software performance, SEE CHART 2]***

**SMP (symmetric multi-processing) scalability** – The availability of standard, non-proprietary servers with more processors than before means that SAS software users on Pentium III Xeon processor-based servers can get more throughput than before. The Pentium III Xeon processor

supports up to eight processors per system, a doubling of the previous standard of four per system. When running software such as SAS/IntrNet Application Server on Pentium III Xeon processor-based servers, a doubling of processing throughput should be possible.

**CHART 1**

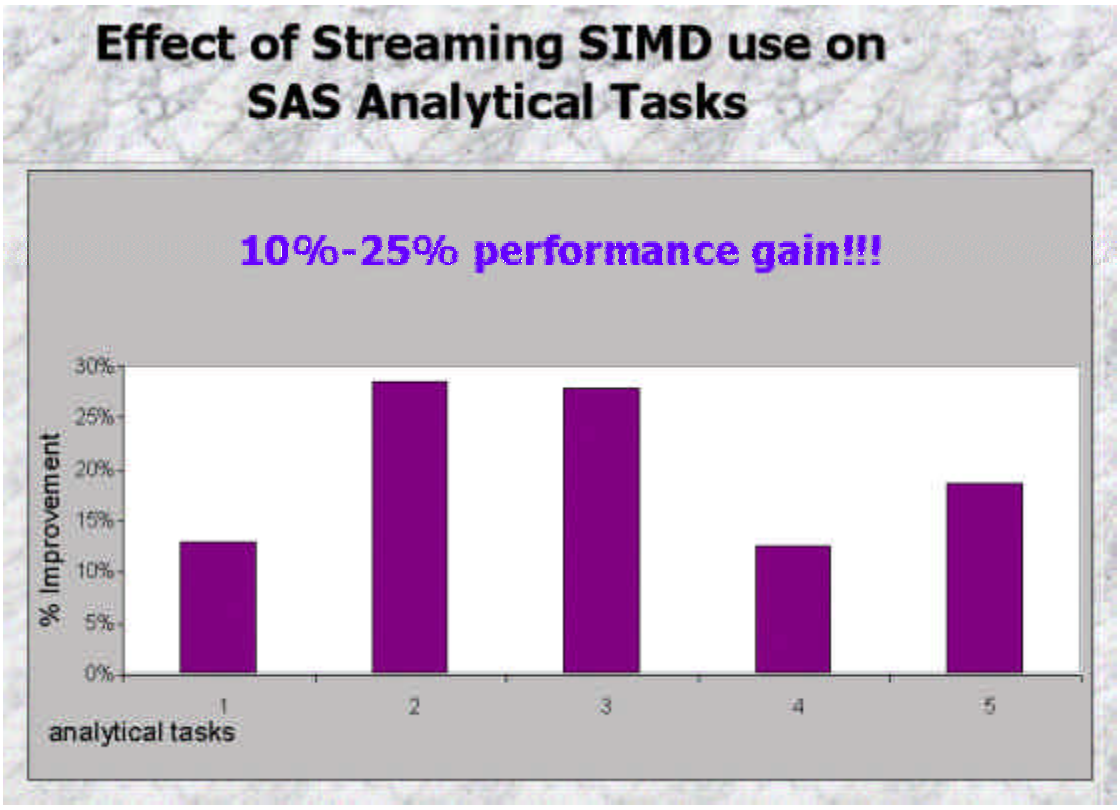



The SAS System for Windows was tested on a server with a 500 MHz Pentium III Xeon processor, running Windows NT 4.0, Enterprise Edition, with the ESMA driver. The SAS dataset used for the test was 3 Gigabytes in size. The elapsed times were measured in two series of tests: SAS without ESMA and SAS with ESMA. The results are shown as a percentage improvement. These were the test cases:

- T1: Sequential Writes
- T2: Sequential Reads
- T3: Random Reads
- T4: Mixture -- Random Reads and Sequential Writes
- T5: Index Creation (Internally Must Sort Data)

- T6: Random Writes

**CHART 2**




The SAS System for Windows was tested server with a 500 MHz Pentium III Xeon processor, running Windows NT 4.0, Enterprise Edition. It was tested before and after adding support for Streaming SIMD extensions instructions into an internal SAS function. The function changed for these tests was a routine that calculates the inner product of two vectors. PROC FACTOR, an important SAS statistical procedure in SAS software used for factor analysis, uses this routine heavily. SAS Institute's data-mining solution, Enterprise Miner software, and many other analytic modeling techniques in SAS software also use this routine. These were the test cases:

- T1: PROC FACTOR run on dataset with 20,000 observations
- T2: PROC FACTOR run on dataset with 50,000 observations
- T3: PROC FACTOR run on dataset with 100,000 observations
- T4: PROC FACTOR run on dataset with 200,000 observations
- T5: PROC FACTOR run on dataset with 300,000 observations

*March 17, 1999*